International Journal for Research in Engineering Application & Management (IJREAM)
ISSN : 2454-9150 Vol-05, Issue-03, June 2019

Role of Hashtag (#) in Social Media and Data Mining: A Review

Vandana Shukla, MCA student, BVIMIT Navi Mumbai India, vandanashukla9895@gmail.com

Nidhi, Asst. Professor, BVIMIT Navi Mumbai India, mca.nidhipoonia@gmail.com

Abstract: With the rapid growth of Social media platform, its significance and usage has also grown so far. Social Media platforms such as Twitter, Instagram, Facebook, etc. are powerful communication and information sharing tools which are used by many people resulting in production of heaps and heaps of random data. With such a rapid growth of data each day, IT communities have always adapted new ways to absorb this random data, finding patterns to provide better communication features within the social media. The technology behind finding patterns among these random data is Data Mining and Text Mining is one of the branches implemented in these social media platforms improving quality of usage of these platforms making them play a prominent role in an average person's life. One such data mining technique is Hashtag. Hashtag is a social media term used on various social media platforms with posts by the users themselves. It is used with the (#) sign following by keywords with no spaces, e.g. #helloworld. Moreover, Hashtag is a data collecting tool which is analyzed to create AI models, Neuron network models and various other data mining techniques which is further applied in different fields as a powerful business tool for different organizations, institutions and social media platforms. In this paper, we are briefly discussing about origin of hashtag, its growth throughout the years, its usage over different social media platforms and its applications as data mining models and in real life

*Keywords — **Data mining, hashtag, role of hashtag, social media, text mining techniques.***

I. INTRODUCTION

As said by Danah M. Boyd and Nicole B. Allison [1], Social media is a place, community or a platform where people can interact, share their opinions, photos, emotions, experience and debate on any topic with their social contacts and also increase their social contacts. Data Mining has been in existence even before the Social Media has walked into an average person's life. For some time, Data mining was used to find various hidden patterns among the random data until it made a breakthrough of generating the heaps of data already in a pattern. Although it was a random idea to implement data clustering using a defined symbol followed by keywords. It gave birth to whole new ways of patterning data before generating them.

Hashtag is a metadata tag used preceded by pound (#) sign on social media platforms such as twitter, Instagram, YouTube, Pinterest, etc. The general idea behind introduction of hashtag was to allow users to apply dynamic, user-generated tagging which allows users to easily identify posts related to the respective terms. Hash-tag is generated by the user using the (#) sign followed by terms or words specified by the user. The styling of hashtag includes hash (#) sign followed by words with no spaces and numeric, e.g., #helloworld, #writingapaper. It should be noted that these terms are not case-sensitive and cannot be started with numeric, e.g., #123donotuseit, whereas numeric can be followed by characters, e.g., #canuse123. The keywords could be of any language, capitalized, mixed, lowercase or uppercase. Hash-tag grouping came into existence by a tweet made by Chris Messina on twitter suggesting the idea of using hash-sign(pound - sign) as easy tagging of posts in order to form a thread of conversation on twitter termed as "group chat". Below figure 1.1, is the image of first tweet about usage of hashtag referred by [15].

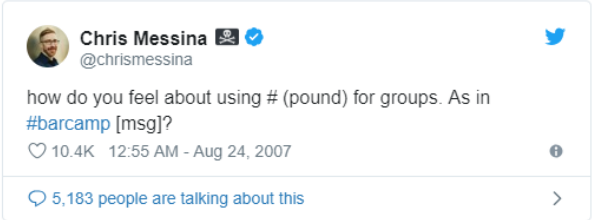

Figure I-1 First tweet about hashtag

This feature of hashtag of allowing users to create their own unique hashtags has fascinated users so much that the usage of hashtags spread like a wildfire and every corporation or social media platform was forced to put a set of rules such as hashtags for a single post should be less than or equal to thirty(30) in number. It also has the ability to grab attention among the users almost immediately and spreading the

 DOI : 10.35291/2454-9150.2019.0191

information as quickly as possible making users aware or participate from almost anywhere in the world and also voice their opinion related to the information or the issue or an event. Usage became so common that it made others companies or social media platforms to adapt them as soon as possible. The usage of Hashtags has influenced various social issues and areas such as Political events, entertainment, corporal marketing or digital marketing, a social cause, communication among other social contacts.

The role of hashtag can be defined in three ways, the first is its application in data mining methods as it serves as data collector for the models. The next role of hashtag is its application in real life across various social media platforms to gain attention by different organizations using the different models generated by collection of data using hashtag. The third important role is its existence on social media platforms benefitting them to both gain attention and generate patterned data.

II. HISTORY

Twitter is an American social media service created by Jack Dorsey, Noah Glass, Biz Stones and Evan Williams in March 2006. It allows its registered users to share photos and videos and post or send messages limited to 140 characters following the Short Message Service format, which was recently doubled to 280 characters known as "tweets". It gained popularity very rapidly and now there are more than 100 million registered users and approximately 340 million tweets are posted every day.

In August 2007, Chris Messina officially introduced the world to Hashtag by posting a tweet asking people how they feel about using pound (#) sign for group conversations. Following is the proposed syntax for the usage of #tag by Messina. Below figure 2.1 represents initial introduction of syntax for the usage of Hashtag on twitter by Chris Messina referred from [14].


Syntax

- `follow #tag` : subscribe to all updates tagged with *#tag*
- `follow username#tag` : subscribe to all updates tagged with *#tag* from a specific user
- `leave #tag` : unsubscribe to a tag; you will still get updates with this tag from your friends
- `leave username#tag` : unsubscribe to a specific from a specific user
- `remove #tag` : completely remove all incoming posts tagged with *#tag*, even from your friends
- `#tag message` : creates a status in the *#tag* channel
- `#tag !message` : creates a status that is only visible to people subscribed to channel tag *#tag*

Figure II-1 Syntax introduced for Hashtag

Messina stated that the idea was influenced by Similar social websites such as Jaiku created by Jyri Engestrom which was influenced by Jarkko Oikarinen's IRC. The usage of hash was simple as they used the sign to name or create channels for group communications followed by a word.

There was a slow start for the usage of hashtags in the beginning but eventually it gained popularity due to a political event in Iran. During the elections 2009-10, people within and out of the country happened to realize the usage of Hashtag and started tweeting about the elections in both English and Persian language in order to protest against presidential election results. This event provided the required limelight on usage of hashtag and people started using it as writing style for posting tweets.

III. DATA MINING APPLICATIONS OF HASHTAG

Text mining is a branch of data mining which is used to discover knowledge or patterns for better or further use. The text mining process undergoes various stages that is, collection of data, then preparing the data in order to analyse further, then applying clustering and classification methods to analyse text, forming a model which defines a pattern resulting in final stage that is knowledge discovery. Following figure 3.1, depicts the various stages of text mining.

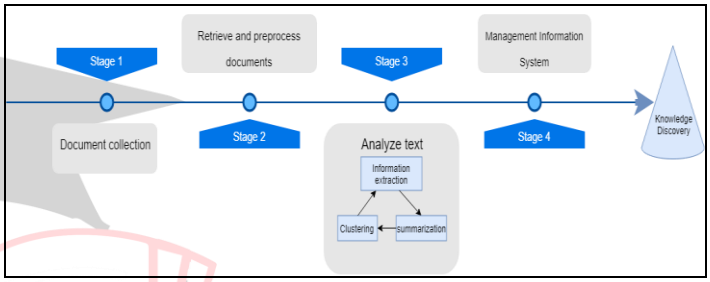

Figure III-1 Text Mining Process

As discussed above that Hashtag is a data collecting tool in a patterned manner it is evident it has been influencing many works for a long time. Following are related works of hashtag explained:

- Boyd, danah; Ellison, Nicole (2008). [1] explained the origin of Social network Sites and proposing a definition, "We define social network sites as web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system. The nature and nomenclature of these connections may vary from site to site."
- Bidisha Samanta, Abir De, Abhijnan Chakraborty, Niloy Ganguly [2] proposed Large Margin Point Process (LMPP), a probabilistic framework for Hashtag propagation integrating two main factors: hashtag-tweet influence and hashtag-hashtag competitions. The proposed model can also accurately predict the relative rankings of competing hashtags providing valuable data for further analysis.
- Jie Liu, Zhicheng He and Yalou Huang [3] devised an Embedding model, Hashtag2Vec, using

 DOI : 10.35291/2454-9150.2019.0191

multiple relations among hashtag-hashtag, hashtag-tweet, tweet-word, word-word based on hierarchical heterogeneous network. It is capable of embedding hashtags and short social texts as well.

- Ramzan Talib, Muhammad Kashif Hanif, Shaeela Ayesha, and Fakeeha Fatima [4] discussed and analysed different text mining techniques, their applications in real life and also their adverse effects.
- Muhammad Asif, Malik Muhammad Saad, Nadeem Akhtar, Hina Asmat, Mujtaba Husnain, Muhammad Asghar [5] performed experiments in order to find out that do hashtags represent the content of the tweet they are used to summarize for? They followed four steps process of data collecting , segregating the compound hashtags , applying different methods to the data like Adaptive Extended Lesk method, Hirst & St-onge, Lin, Wu & Palmer, Path Length, Resnik, leacock & chodorow, gloss vectors (pair-wise) ,gloss vectors (vectors), jiang & conrath and finally normalizing all the obtained results in order to compare.
- Jianjun YU, Tongyu ZHU [6] introduced the hashtag recommender system using three features of user interest that is, hashtag textual information, user behaviour and time. The introduced models are a linear combined model and an enhanced session-based temporal graph model.
- Van den Berg, J.A., 2014 [7] briefed about existence and origin of hashtag based on social media platform twitter and continued to brief various hashtag led events and their effects.
- Jun Ma , Chong Feng, Ge Shi, Xuewen Shi, Heyang Huang [8] proposed long short-term memory based model to recommend reliable hashtags to the user. It uses temporal enhanced selective sentence-level attention to reduce the influence of wrong labelled microblogs to the classifier. Experimental results using a dataset of 1.7 million microblogs collected from SINA Weibo microblogs demonstrated that the proposed method could achieve significantly better performance than the state-of-the-art methods.
- Kavinga Yapa Abeywardana, Ginige A.R., Herath N., Somarathne H.P., Thennakoon T.M.N.S [9] proposed an application for hashtag generation and content authenticator by analysing text and image using prediction models to classify and filter the hashtags for the user. Following figure 3.2, is the system architecture diagram of the proposed application.

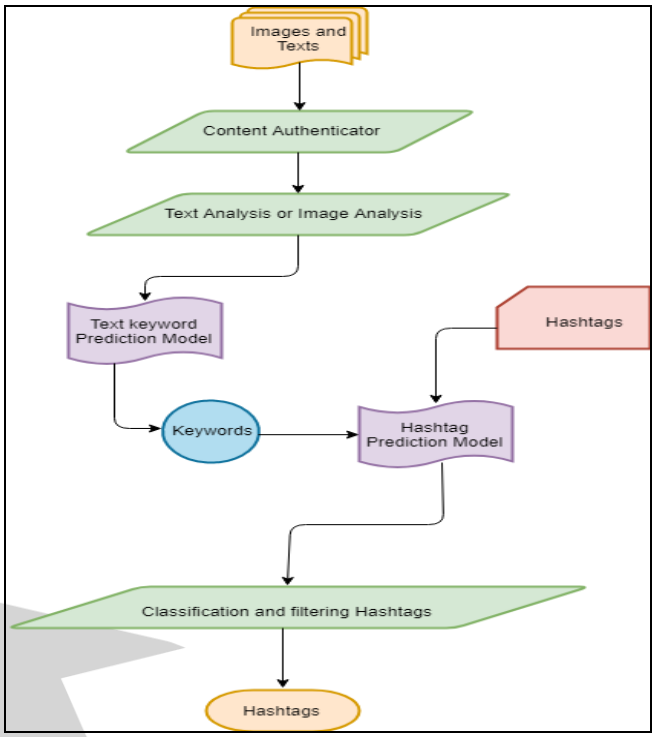

Figure III-2 System Architecture Structure

IV. REAL LIFE APPLICATIONS OF HASHTAG

As said before, with the increasing use of hashtag as style of posting on various social media platforms it has greatly impacted or also can be said to have changed the usage of social media in present days. Following are some of the factors that point the impact caused by this Meta tagging tool.

A. Politics

In Politics, There has been various instances in history where social media has greatly impacted politics. For instance in 2008, former U.S. president, Mr. Barack Obama tactfully utilized corporation of hashtags in twitter as most congressmen were users of twitter compared to his rival party. He carefully used the potential of social media by advertising his visions, thoughts for people of America and his past achievements throughout his work history which made people aware of his capability resulting in him winning the elections.

Another such instance is a recent political event in Indian Lok Sabha elections 2019 where a snide comment made by opposing party on current ruling party, resulted in a controversial hashtag, #meinbhichowkidar increased the attention towards ruling party winning them the elections with a clean sweep. The point to be noted here is that this came as an adverse effect on the public image of opposing party and a small political remark determined the results of Indian Lok Sabha elections 2019.

 DOI : 10.35291/2454-9150.2019.0191

B. Business Tool

With the gaining popularity of social media, now there are many companies that are run using social media. Hence, hashtag plays a key role in branding of these companies. Usage of hashtags expands one's content reach, gradually increases the target audience, gathers attention required to the company's content and also improves the SEO. The usage of hashtags are based upon three factors that are Brand and Campaign specific, Trending and Content. In the first kind, Companies are required to use hashtags that represent their motto or as said their tagline. For example, KitKat using #takeabreak, Coca Cola using #shareacoke, KFC using #NationalFriedChickenDay are the examples for a successful marketing strategies who engaged their audiences into their marketing of brands. On the other side, the greatest disadvantage of social media is its unpredictable nature and chaining action, one great example for an unsuccessful marketing is about McDonalds, the multinational food chain company initiated a hashtag as a part of marketing, #McDstories. Although, the motive behind this hashtag was a general idea of people sharing their best memories in McDonald's stores backfired very spectacularly when people started posting about their bad experiences in McDonald's franchise. This grabbed a great deal of attention to the company which was negative resulting in a considering amount of loss.

Everyone is now aware that the core nature of social media is very differing due to its prime feature of spreading the word with lightning speed a new term has been introduced to define the popularity one gains with a single post that is, "Trending" or "Viral". These terms are referred to the posts that have gained maximum amount of attention in the least amount of time. The usage of hashtags has become frequent that one google search as top trending hashtags and can get lists in results that are updated on a daily basis. Now, Content-driven hashtags are not trending or popular ones but are commonly used terms that are relatable to most people. Content driven hashtags can be generally referred to different products such as #Latte, lifestyle driven like #mondayblues or event based like #donateblood or location based such as #manali or #londonbridge.

C. Social Campaigns

Social media has played a very important role in providing more than the required attention to the campaigns be it in any given field such as in health awareness there has been various campaigns about HIV / AIDS, cancers, Diabetes, Migraines, Vaccinations, self-care, diets and many more. In Entertainment field, there has been #metoo campaign that has run for a long time in Indian entertainment industry where women shared their disturbing experiences while pursuing their career or in their career span. Many famous celebrities were accused of harassing the women working under them which brewed a great deal of controversy throughout the nation. Originated from the Hollywood, the first tweet was posted by an American actress Alyssa Milano shared her story of being harassed on camera for a rape scene for a movie by the director Harvey Weinstein in October 2017. This tweet went viral so quickly by people responding by sharing their stories, not only women but also men who suffered sexual harassment or assault in their lives. It is estimated that #metoo was used more than 19 million times on twitter itself. One can guess its usage on other platforms. One of the many reasons for the popularity of this disruptive campaign is that it affected people psychologically, positive for the victims who shared their stories, mentally relieved them from the burdens they carried for so long and negative to those who were accused of harassment or assault by the victims in their stories causing them lose their social dignity in the society.

In Data Mining, hashtag became a data collecting tool upgrading works like natural language processing, sentimental analysis, prediction models, AI models, hazard management models, micro blogging websites, recommender systems, and etcetera.

As observed above, hashtag has been in existence for more than a decade now and it has continuously revolutionizing the IT industry and a common person's daily life. There have been many instances where it was removed by social media websites but its psychological effect on people's mind kept it in existence .It can be seen that the growth of hashtag is growing with the social media. And it will continue to help grow and better the data mining techniques or works.

V. Usage of Hashtag on Various Social Media Websites

As twitter was rapidly gaining popularity because of this new writing style of posting. This lead the rival social media services to adapt this Meta tag technique into their algorithms. As stated in Wikipedia [13], In January 2011, Instagram introduced its users to hashtags helping them to discover photos, videos, posts and their social contacts. It encouraged its users to make both specific and relevant tags in order to make their posts stand out. In June 2013, Facebook adapted Hashtag in to their algorithm in order to provide similar features in their social service. Which lead YouTube to adapt hashtag in February 2016 in order to allow users to find related videos online. Then again in September 2017, Pinterest introduced hashtag on their website in order to provide the similar features to their users.

 DOI : 10.35291/2454-9150.2019.0191

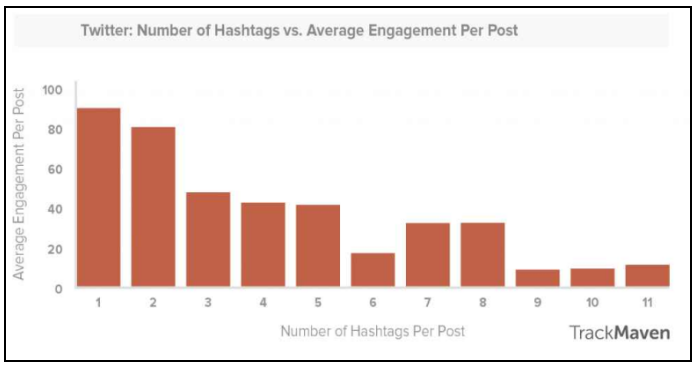

Figure V-1 Graph on engagement per post on Twitter

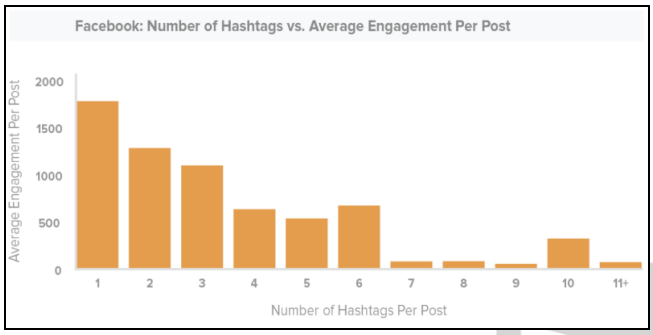

Figure V-2 Graph on Engagement per post on Facebook

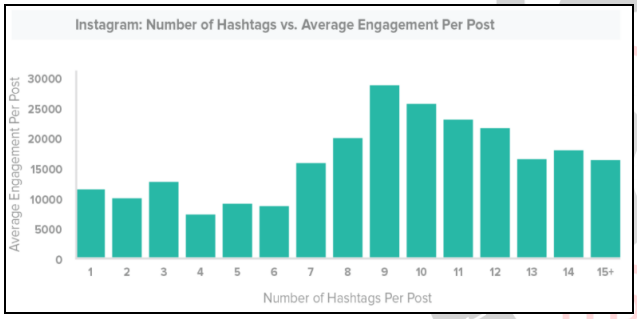

Figure V-3 graph on engagement per post on Instagram.

The above figures 5.1, 5.2, 5.3 are the graphs of three different social websites giants representing the engagement of users per post with the number of hashtags used with that post, referred from the work of Maddy Osman[16]. As observed, it can be seen that in the graph for Facebook and twitter, the increase in number of hashtags per post decreases the engagement of users on that post whereas in the graph of Instagram, the scenario is completely different. The increase in number of hashtags per post increases the number of engaging users for that post. As compared, the effect of engagement of user differs on each websites with different numbers of usage of hashtags.

VI. CONCLUSION

As discussed in this paper, we can conclude that the most fascinating fact about Hashtag is that it was introduced to the world as a simple data collecting tool which surprisingly revolutionized text mining techniques and affected people positively with its simplicity in use and making an impact in huge and also has been in existence for more than a decade now. We can collectively agree to Van den Berg[7], who stated that *" The popularity of hashtag lies within its functional usage and the functionality of this symbol is found in the common practice of sorting and selecting thematically related information from a torrent of messages within the context of social media platforms"*. We hope that the developments brought by Hashtag in data mining industry and people's lives continue to grow leading to better future of IT.

REFERENCES

[1] Jianjun YU, Tongyu ZHU: "Combining long-term and short-term user interest for personalized hashtag recommendation". Front. Comput. Sci.,2015, 9(4): 608–622 DOI 10.1007/s11704-015-4284-x.

[2] Bidisha Samanta, Abir De, Abhijnan Chakraborty, Niloy Ganguly:" LMPP: A Large Margin Point Process Combining Reinforcement and Competition for Modeling Hashtag Popularity", Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI-17).

[3] Jie Liu, Zhicheng He and Yalou Huang:" Hashtag2Vec: Learning Hashtag Representation with Relational Hierarchical Embedding Model", Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI-18).

[4] Ramzan Talib, Muhammad Kashif Hanif, Shaeela Ayesha, and Fakeeha Fatima:" Text Mining: Techniques, Applications and Issues", (IJACSA) International Journal of Advanced Computer Science and Applications,Vol. 7 No. 11, 2016.

[5] Muhammad Asif, Malik Muhammad Saad, Nadeem Akhtar, Hina Asmat, Mujtaba Husnain, Muhammad Asghar : "Hashtag the Tweets: Experimental Evaluation of Semantic Relatedness Measures" , (IJACSA) International Journal of Advanced Computer Science and Applications,Vol. 7, No. 6, 2016.

[6] Boyd, danah; Ellison, Nicole (2008). "Social Network Sites: Definition, History, and Scholarship". Journal of Computer-Mediated Communication 13: 210–230.doi:10.1111/j.1083-6101.2007.00393.x

[7] Van den Berg, J.A., 2014,'The story of the hashtag(#):A practical theological tracing of the hashtag(#) symbol on Twitter', HTS Teologiese Studies/Theological Studies 70(1), Art. #2706, 6 pages. http://dx.doi.org/10.4102/hts.v70i1.2706.

[8] Jun Ma , Chong Feng, Ge Shi, Xuewen Shi, Heyang Huang: "Temporal enhanced sentence-level attention model for hashtag recommendation", ISSN 2468-2322 ,21st January 2018, doi: 10.1049/trit.2018.0012,www.ietdl.org.

 DOI : 10.35291/2454-9150.2019.0191

[9] Kavinga Yapa Abeywardana, Ginige A.R., Herath N., Somarathne H.P., Thennakoon T.M.N.S. :" Hashtag Generator and Content Authenticator", (IJACSA) International Journal of Advanced Computer Science and Applications,Vol. 9, No. 9, 2018.

[10] article: George N, Britto DR, Krishnan V, Dass LM, Prasant HA, Aravindhan V.Assessment of hashtag (#) campaigns aimed at health awareness in social media. J Edu Health Promot 2018;7:114.

[11] Davis, Bud (2013) "Hashtag Politics: The Polyphonic Revolution of #Twitter," Pepperdine Journal of Communication Research: Vol. 1,Article 4. Available at: http://digitalcommons.pepperdine.edu/pjcr/vol1/iss1/4.

[12] Yong Nahm, Un & J. Mooney, Raymond. (2002). Text Mining with Information Extraction.

[13] Hashtag from Wikipedia, https://en.wikipedia.org/wiki/Hashtag.

[14] The Surprising History of Twitter's Hashtag Origin and 4 Ways to Get the Most out of Them by Belle Beth Cooper https://buffer.com/resources/a-concise-history-of-twitter-hashtags-and-how-you-should-use-them-properly .

[15] Andrea Sandre , "Twitter didn't invent the hashtag… Chris Messina did!" , https://hackernoon.com/twitter-didnt-invent-the-hashtag-chris-messina-did-1020969abfcd.

[16] Maddy Osman, july 11 2017, "How to Use Hashtags on Every Social Media Network", https://sproutsocial.com/insights/how-to-use-hashtags/

 DOI : 10.35291/2454-9150.2019.0191